16/pats

The invention relates to polynucleotides coding for a fucosyl transferase. Furthermore, the invention relates to partial sequences of these polynucleotides as well as to vectors comprising these polynucleotides, recombinant host cells, plants and insects transfected with the polynucleotides or with DNA derived therefrom, respectively, as well as to glycoproteins produced in these systems.

Glycoproteins exhibit a variety and complexity of carbo-hydrate units, the composition and arrangement of the carbohydrates being characteristic of different organisms. The oligosaccharide units of the glycoproteins have a number of tasks, e.g. they are important in regulating metabolism, they are involved in transmitting cell-cell interactions, they determine the circulation periods of proteins in circulation, and they are decisive for recognizing epitopes in antigen-antibody reactions.

The glycosylation of glycoproteins starts in the endo-plasmatic reticulum (ER), where the oligosaccharides are either bound to asparagine side chains by N-glycosidic bonds or to serine or threonine side chains by O-glycosidic bonds. The N-bound oligosaccharides contain a common core from a penta-saccharide unit which consists of three mannose and two N-acetyl glucose amine residues. To further modify the carbohydrate units, the proteins are transported from the ER to the Golgi complex. The structure of the N-bound oligosaccharide units of glycoproteins is determined by their conformation and by the composition of the glycosyl transferases of the Golgi compartments in which they are processed.

It has been shown that the core pentasaccharide unit in the Golgi complex of some plant and insect cells is substituted by xylose and α1,3-bound fucose (P. Lerouge et al., 1998, Plant Mol. Biol. 38, 31-48; Rayon et al., 1998, L. Exp. Bot. 49, 1463-1472). The heptasaccharide "MMXF^3" forming constitutes the main oligosaccharide type in plants (Kurosaka et al., 1991, J. Biol. Chem., 266, 4168-4172). Thus, e.g., the horseradish peroxidase, carrot β-fructosidase and Erythrina cristagalli comprise lectin as well as the honeybee venom phospholipase A2 or the neuronal membrane glycoproteins from insect embryos α1,3-fucose residues which are bound to the glycan core. These structures are also termed complex N-glycans or mannose-deficient or truncated N-glycans, respectively. The α-mannosyl residues may be further replaced by

GlcNAc, to which galactose and fucose are bound so that a structure is prepared which corresponds to the human Lewis a-epitope (Melo et al., 1997, FEBS Lett 415, 186-191; Fitchette-Laine et al., 1997, Plant J. 12, 1411-1417).

Neither xylose nor the α1,3-bound fucose exist in mammalian glycoproteins. It has been found that the core-α1,3-fucose plays an important role in the epitope recognition of antibodies which are directed against plant and insect N-bound oligosaccharides (I.B.H. Wilson et al., Glycobiology Vol. 8, No. 7, pp. 651-661, 1998), and thereby trigger immune reactions in human or animal bodies against these oligosaccharides. The α1,3-fucose residue furthermore seems to be one of the main causes for the widespread allergic cross reactivity between various plant and insect allergens (Tretter et al., Int. Arch. Allergy Immunol. 1993; 102:259-266) and is also termed "cross-reactive carbohydrate determinant" (CCD). In a study of epitopes of tomatoes and grass pollen, also α1,3-bound fucose residues were found as a common determinant, which seems to be the reason why tomato and grass pollen allergies frequently occur together in patients (Petersen et al., 1996, J. Allergy Clin. Immunol., Vol. 98, 4; 805-814). Due to the frequent occurrence of immunological cross reactions, the CCDs moreover mask allergy diagnoses.

The immunological reactions triggered in the human body by plant proteins are the main problem in the medicinal use of recombinant human proteins produced in plants. To circumvent this problem, α1,3-core-fucosylation would have to be prevented. In a study it could be demonstrated that oligosaccharides comprising an L-galactose instead of an L-fucose (6-deoxy-L-galactose) nevertheless are biologically fully active (E. Zablackis et al., 1996, Science, Vol. 272). According to another study, a mutant of the plant Arabidopsis thaliana was isolated in which the N-acetyl-glucosaminyl transferase I, the first enzyme in the biosynthesis of complex glycans, is missing. The biosynthesis of the complex glycoproteins in this mutant thus is disturbed. Nevertheless, these mutant plants are capable of developing normally under certain conditions (A. Schaewen et al, 1993, Plant Physiol. 102; 1109-1118).

To purposefully block the binding of the core-α1,3-fucose in an oligosaccharide without also interfering in other glycosylation steps, merely that enzyme would have to be inactivated which

is directly responsible for this specific glycosylation, i.e. the core-α1,3-fucosyl transferase. It has been isolated and characterized for the first time from mung beans, and it has been found that the activity of this enzyme depends on the presence of non-reducing GlcNAc ends (Staudacher et al., 1995, Glycoconjugate J. 12, 780-786). This transferase which only occurs in plants and insect, yet not in human beings or in other vertebrates, would have to be inactivated on purpose or suppressed so that human proteins which are produced in plants or in plant cells or also in insects or in insect cells, respectively, do no longer comprise this immune-reaction-triggering epitope, as has been the case so far.

The publication by John M. Burke "Clearing the way for ribozymes" (Nature Biotechnology 15:414-415; 1997) relates to the general mode of function of ribozymes.

The publication by Pooga et al., "Cell penetrating PNA constructs regulate galanin receptor levels and modify pain transmission in vivo" (Nature Biotechnology 16:857-861; 1998) relates to PNA molecules in general and specifically to a PNA molecule that is complementary to human galanin receptor type 1 mRNA.

US 5,272,066 A relates to a method of changing eukaryotic and prokaryotic proteins to prolongue their circulation in vivo. In this instance, the bound oligosaccharides are changed with the help of various enzymes, among them also GlcNAc-α1→3(4)-fucosyl transferase.

EP 0 643 132 A1 relates to the cloning of an α1,3-fucosyl transferase isolated from human cells (THP-1). The carbohydrate chains described in this publication correspond to human sialyl Lewis x- and sialyl Lewis a-oligosaccharides. The specificity of the enzyme from human cells is quite different than that of fucosyltransferase from plant cells.

It is an object of the present invention to clone and to sequence the gene which codes for a plant fucosyl transferase, and to prepare vectors comprising this gene, DNA fragments thereof or an altered DNA or a DNA derived therefrom, to transfect plants and insects as well as cells thereof with one of these vectors, to produce glycoproteins that do not comprise the normally occurring α1,3-core-fucose, as well as to provide corresponding methods therefor.

The object according to the invention is achieved by a DNA

molecule comprising a sequence according to SEQ ID NO: 1 (in this disclosure also the IUPAC code has been used, "N" meaning inosin) with an open reading frame from base pair 211 to base pair 1740 or being at least 50% homologous to the above sequence or hybridizing with the above-indicated sequence under stringent conditions, or comprising a sequence which has degenerated to the above DNA sequence due to the genetic code, the sequence coding for a plant protein which has fucosyl transferase activity or is complementary thereto.

This sequence which has not been described before can be perfectly used for any experiments, analysis and methods for production etc. which relate to the plant fucosyl transferase activity. Here the DNA sequence as well as the protein coded by this sequence are of interest. However, in particular the DNA sequence will be used for the inhibition of the fucosyl transferase activity.

The open reading frame of the SEQ ID NO: 1 codes for a protein with 510 amino acids and with a theoretical molecular weight of 56.8 kDa, a transmembrane portion presumably being present in the region between Asn36 and Gly54. The calculated pI value of the encoded protein of the sequence according to SEQ ID NO: 1 is 7.51.

The activity of the plant fucosyl transferase is detected by a method and measured, the fucosyl transferase being added to a sample comprising labelled fucose and an acceptor (e.g. a glycoprotein) bound to a carrier, e.g. Sepharose. After the reaction time, the sample is washed, and the content of bound fucose is measured. The activity of the fucosyl transferase in this case is seen as positive if the activity measurement is higher by at least 10 to 20%, in particular at least 30 to 50%, than the activity measurement of the negative control. The structure of the glycoprotein may additionally be verified by means of HPLC. Such protocols are prior art (Staudacher et al. 1998, Anal. Biochem. 246, 96-101; Staudacher et al. 1991, Eur. J. Biochem. 199, 745-751).

For example, fucosyl transferase is admixed to a sample comprising radioactively labelled fucose and an acceptor, e.g. GlcNAcβ1-2Manα1-3(GlcNAβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn. After the reaction time, the sample is purified by anion exchange chromatography, and the content of bound fucose is measured.

From the difference of the measured radioactivity of the sample with acceptor and that of a negative control without acceptor, the activity can be calculated. The activity of the fucosyl transferase is already evaluated as positive if the radioactivity measured is at least 30-40% higher than the measured radioactivity of the negative sample.

The pairing of two DNA molecules can be changed by selection of the temperature and ionic strength of the sample. By stringent conditions, according to the invention conditions are understood which allow for an exact, stringent, binding. For instance, the DNA molecules are hybridized in 7% sodium dodecyl sulfate (SDS), 0.5M NaPO4, pH 7.0, 1mM EDTA at 50°C, and washed with 1% SDS at 42°C.

Whether sequences have an at least 50% homology to SEQ ID NO: 1 can be determined e.g. by means of the program FastDB of EMBL or SWISSPROT data bank.

Preferably, the sequence of the DNA molecule of the invention encodes a protein with a GlcNAc-α1,3-fucosyl transferase activity, in particular with a core-α1,3-fucosyl transferase activity.

As described above the core of α1,3-fucosyl transferase is present in plants and insects, however, not in the human body, so that in particular this DNA sequence is useful in analysis and experiments as well as methods for production which are fucosyl transferase specific.

By a core-α1,3-fucosyl transferase, in particular GDP-L-Fuc:Asn-bound GlcNAc-α1,3-fucosyl transferase is understood. Within the scope of the present invention, the term α1,3-fucosyl transferase as a rule particularly means core-α1,3 fucosyl transferase. For the above-described activity measurement, in particular acceptors having a non-reducing GlcNAc terminus are used. Such acceptors are, e.g., GlcNAcβ1-2Manα1-3(GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn, GlcNAcβ1-2Manα1-3(GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAcβ1-Asn and GlcNAcβ1-2Manα1-3[Manα1-3(Manα1-6)Manα1-6]Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn. Whether the fucose is bound or not can furthermore be determined by measuring the insensitivity relative to N-glycosidase F, which can be detected by means of mass spectrometry.

Preferably, the DNA molecule according to the invention comprises at least 70-80%, particularly preferred at least 95%, ho-

mology to the sequence according to SEQ ID NO: 1. This sequence codes for a particularly active GlcNAc-α1,3-fucosyl transferase.

Since the DNA sequence can be more or less changed according to the plant or the insect a sequence which shows, for example, 70 % homology to a sequence according to SEQ ID No 1 has also a fucosyl transferase activity which is sufficient in order to be used in analysis, experiments or methods of production as above described.

According to a further advantageous embodiment, the DNA molecule comprises 2150 to 2250, in particular 2198, base pairs. This DNA molecule comprises 100 to 300, preferably 210, base pairs upstream in front of the start codon, as well as 350 to 440, in particular 458, base pairs downstream after the stop codon of the open reading frame, wherein the end of the DNA molecule preferably comprises a 3'-poly(A)-tail. In this manner, a faultless regulation on translation level is ensured and a DNA molecule is provided which is particularly efficient and unproblematic for the coding of an active GlcNAc-α1,3-fucosyl transferase.

The present invention moreover relates to a DNA molecule which comprises a sequence according to SEQ ID NO: 3 or comprising a sequence having at least 85%, particularly preferred at least 95%, in particular at least 99%, homology to the above-identified sequence or which, under stringent conditions, hybridizes with the above-indicated sequence or which has degenerated to the above-indicated DNA sequence due to the genetic code. The homology preferably is determined with a program which recognizes insertions and deletions and which does not consider these in the homology calculation. This nucleotide sequence codes for a conserved peptide motif, which means that the plurality of the active and functioning GlcNAc-α1,3-fucosyl transferases comprises the amino acid sequence encoded thereby. In this instance, the sequence may either have the same size as the sequence according to SEQ ID NO: 3, or, of course, it may also be larger. This sequence has a smaller length than the sequence which codes the complete protein and is therefore less sensitive with respect to recombination, deletion, or any other mutations. Due to the conservative motif and its higher stability this sequence is particularly advantageous for sequence recognising test.

SEQ ID NO: 3 comprises the following sequence:

5'-GAAGCCCTGAAGCACTACAAATTTAGCTTAGCGTTTGAAAATTCGAATGAGGAAG
ATTATGTAACTGAAAAATTCTTCCAATCCCTTGTTGCTGGAACTGTCCCT-3'

In a further aspect, the present invention relates to a DNA molecule which comprises a partial sequence of one of the above-indicated DNA molecules and has a size of from 20 to 200, preferably from 30 to 50, base pairs. The DNA molecule may, e.g., be utilized to bind, as a probe, to complementary sequences of GlcNAc-α1,3-fucosyl transferases so that they can be selected from a sample. In this manner, further GlcNAc-α1,3-fucosyl transferases from the most varying plants and insects can be selected, isolated and characterized. Any desired one or also several different partial sequences may be used, in particular a part of the conserved motif already described above.

In doing so, it is particularly advantageous if one of the above-indicated DNA molecules is covalently associated with a detectable labelling substance. As the labelling substance, any common marker can be used, such as, e.g., fluorescent, luminescent, radioactive markers, non-isotopic markers, such as biotin, etc. In this manner, reagents are provided which are suitable for the detection, selection and quantitation of corresponding DNA molecules in solid tissue samples (e.g. from plants) or also in liquid samples, by means of hybridizing methods.

A further aspect of the invention relates to a biologically functional vector which comprises one of the above-indicated DNA molecules or parts thereof of differing lengths with at least 20 base pairs. For transfection into host cells, an independent vector capable of amplification is necessary, wherein, depending on the host cell, transfection mechanism, task and size of the DNA molecule, a suitable vector can be used. Since a large number of different vectors is known, an enumeration thereof would go beyond the limits of the present application and therefore is done without here, particularly since the vectors are very well known to the skilled artisan (as regards the vectors as well as all the techniques and terms used in this specification which are known to the skilled artisan, cf. also Sambrook Maniatis). Ideally, the vector has a small molecule mass and should comprise selectable genes so as to lead to an easily recognizable phenotype in a cell so thus enable an easy selection of vector-containing and vector-free host cells. To obtain a high yield of DNA and corresponding gene products, the vector should comprise a

strong promoter, as well as an enhancer, gene amplification signals and regulator sequences. For an autonomous replication of the vector, furthermore, a replication origin is important. Polyadenylation sites are responsible for correct processing of the mRNA and splice signals for the RNA transcripts. If phages, viruses or virus particles are used as the vectors, packaging signals will control the packaging of the vector DNA. For instance, for transcription in plants, Ti plasmids are suitable, and for transcription in insect cells, baculoviruses, and in insects, respectively, transposons, such as the P element.

If the above-described inventive vector is inserted into a plant or into a plant cell, a post-transcriptional suppression of the gene expression of the endogenous α1,3-fucosyl transferase gene is attained by transcription of a transgene homologous thereto or of parts thereof, in sense orientation. For this sense technique, furthermore, reference is made to the publications by Baucombe 1996, Plant. Mol. Biol., 9:373-382, and Brigneti et al., 1998, EMBO J. 17:6739-6746. This strategy of "gene silencing" is an effective way of suppressing the expression of the α1,3-fucosyl transferase gene, cf. also Waterhouse et al., 1998, Proc. Natl. Acad. Sci. USA, 95:13959-13964.

Furthermore, the invention relates to a biologically functional vector comprising a DNA molecule according to one of the above-described embodiments, or parts thereof of differing lengths in reverse orientation to the promoter. If this vector is transfected in a host cell, an "antisense mRNA" will be read which is complementary to the mRNA of the GlcNAc-α1,3-fucosyl transferase and complexes the latter. This bond will either hinder correct processing, transportation, stability or, by preventing ribosome annealing, it will hinder translation and thus the normal gene expression of the GlcNAc-α1,3-fucosyl transferase.

Although the entire sequence of the DNA molecule could be inserted into the vector, partial sequences thereof because of their smaller size may be advantageous for certain purposes. With the antisense aspect, e.g., it is important that the DNA molecule is large enough to form a sufficiently large antisense mRNA which will bind to the transferase mRNA. A suitable antisense RNA molecule comprises, e.g., from 50 to 200 nucleotides since many of the known, naturally occurring antisense RNA molecules comprise approximately 100 nucleotides.

For a particularly effective inhibition of the expression of an active α1,3-fucosyl transferase, a combination of the sense technique and the antisense technique is suitable (Waterhouse et al., 1998, Proc. Natl. Acad. Sci., USA, 95:13959-13964).

Advantageously, rapidly hybridizing RNA molecules are used. The efficiency of antisense RNA molecules which have a size of more than 50 nucleotides will depend on the annealing kinetics in vitro. Thus, e.g., rapidly annealing antisense RNA molecules exhibit a greater inhibition of protein expression than slowly hybridizing RNA molecules (Wagner et al., 1994, Annu. Rev. Microbiol., 48:713-742; Rittner et al., 1993, Nucl. Acids Res., 21:1381-1387). Such rapidly hybridizing antisense RNA molecules particularly comprise a large number of external bases (free ends and connecting sequences), a large number of structural subdomains (components) as well as a low degree of loops (Patzel et al. 1998; Nature Biotechnology, 16; 64-68). The hypothetical secondary structures of the antisense RNA molecule may, e.g., be determined by aid of a computer program, according to which a suitable antisense RNA DNA sequence is chosen.

Different sequence regions of the DNA molecule may be inserted into the vector. One possibility consists, e.g., in inserting into the vector only that part which is responsible for ribosome annealing. Blocking in this region of the mRNA will suffice to stop the entire translation. A particularly high efficiency of the antisense molecules also results for the 5'- and 3'-nontranslated regions of the gene.

Preferably, the DNA molecule according to the invention includes a sequence which comprises a deletion, insertion and/or substitution mutation. The number of mutant nucleotides is variable and varies from a single one to several deleted, inserted or substituted nucleotides. It is also possible that the reading frame is shifted by the mutation. In such a "knock-out gene" it is merely important that the expression of a GlcNAc-α1,3-fucosyl transferase is disturbed, and the formation of an active, functional enzyme is prevented. In doing so, the site of the mutation is variable, as long as expression of an enzymatically active protein is prevented. Preferably, the mutation in the catalytic region of the enzyme which is located in the C-terminal region. The method of inserting mutations in DNA sequences are well known to the skilled artisan, and therefore the various possibilities

of mutageneses need not be discussed here in detail. Coincidental mutageneses as well as, in particular, directed mutageneses, e.g. the site-directed mutagenesis, oligonucleotide-controlled mutagenesis or mutageneses by aid of restriction enzymes may be employed in this instance.

The invention further provides a DNA molecule which codes for a ribozyme which comprises two sequence portions of at least 10 to 15 base pairs each, which are complementary to sequence portions of an inventive DNA molecule as described above so that the ribozyme complexes and cleaves the mRNA which is transcribed from a natural GlcNAc-α1,3-fucosyl transferase DNA molecule. The ribozyme will recognized the mRNA of the GlcNAc-α1,3-fucosyl transferase by complementary base pairing with the mRNA. Subsequently, the ribozyme will cleave and destroy the RNA in a sequence-specific manner, before the enzyme is translated. After dissociation from the cleaved substrate, the ribozyme will repeatedly hybridize with RNA molecules and act as specific endonuclease. In general, ribozymes may specifically be produced for inactivation of a certain mRNA, even if not the entire DNA sequence which codes for the protein is known. Ribozymes are particularly efficient if the ribosomes move slowly along the mRNA. In that case it is easier for the ribozyme to find a ribosome-free site on the mRNA. For this reason, slow ribosome mutants are also suitable as a system for ribozymes (J. Burke, 1997, Nature Biotechnology; 15, 414-415). This DNA molecule is particularly advantageous for the downregulation and inhibition, respectively, of the expression of plant GlcNAc-α1,3-fucosyl transferases.

One possible way is also to use a varied form of a ribozmye, i.e. a minizyme. Minizymes are efficient particularly for cleaving larger mRNA molecules. A minizyme is a hammer head ribozyme which has a short oligonucleotide linker instead of the stem/loop II. Dimer-minizymes are particularly efficient (Kuwabara et al., 1998, Nature Biotechnology, 16; 961-965). Consequently, the invention also relates to a biologically functional vector which comprises one of the two last-mentioned DNA molecules (mutation or ribozyme-DNA molecule). What has been said above regarding vectors also applies in this instance. Such a vector can be, for example, inserted into a microorganism and can be used for the production of high concentrations of the above described DNA molecules. Furthermore such a vector is particu-

larly good for the insertion of a specific DNA molecule into a plant or an insect organism in order to downregulate or completely inhibit the GlcNAc-α1,3-fucosyl transferase production in this organism.

According to the invention, there is provided a method of preparing a cDNA comprising the DNA molecule of the invention, wherein RNA is isolted from an insect or plant cell, in particular from hypokotyl cells, by means of which a reverse transcription is carried out after having admixed a reverse transcriptase and primers. The individual steps of this method are carried out according to protocols known per se. For the reverse transcription, on the one hand, it is possible to produce the cDNA of the entire mRNA with the help of oligo(dT) primers, and only then to carry out a PCR by means of selected primers so as to prepare DNA molecules comprising the GlcNAc-α1,3-fucosyl transferase gene. On the other hand, the selected primers may directly be used for the reverse transcription so as to obtain short, specific cDNA. The suitable primers may be prepared e.g. synthetically according to the pattern of cDNA sequences of the transferase. With the help of this method big quantities of the inventive cDNA molecules can be produced quickly in a simple way and with few mistakes.

The invention furthermore relates to a method of cloning a GlcNAc-α1,3-fucosyl transferase, characterized in that the DNA molecule of the invention is cloned into a vector which subsequently is transfected into a host cell or host, respectively, wherein, by selection and amplification of transfected host cells, cell lines are obtained which express the active GlcNac-α1,3-fucosyl transferase. The DNA molecule is inserted into the vector by aid of restriction endonucleases, e.g.. For the vector, there applies what has already been said above. What is important in this method is that an efficient host-vector system is chosen. To obtain an active enzyme, eukaryotic host cells are particularly suitable. One possible way is to transfect the vector in insect cells. In doing so, in particular an insect virus would have to be used as vector, such as, e.g., baculovirus.

Of course, human or other vertebrate cells can also be transfected, in which case the latter would express an enzyme foreign to them.

Preferably, a method of preparing recombinant host cells, in

particular plant or insect cells, or plants or insects, respectively, with a suppressed or completely stopped GlcNac-α1,3-fucosyl transferase production is provided, which is characterized in that at least one of the vectors according to the invention, i.e. that one comprising the inventive DNA molecule, the mutant DNA molecule or the DNA molecule coding for ribozymes or the one comprising the DNA molecule in inverse orientation to the promoter, is inserted into the host cell or plant or into the insect. What has been said above for the transfection also is applicable in this case.

As the host cells, plant cells may, e.g., be used, wherein, e.g., the Ti plasmid with the agrobacterium system is eligible. With the agrobacterium system it is possible to transfect a plant directly: agrobacteria cause root stem galls inplants. If agrobacteria infect an injured plant, the bacteria themselves do not get into the plant, but they insert the recombinant DNA portion, the so-called T-DNA, from the annular, extra chromosomal, tumour-inducing Ti-plasmid into the plant cells. The T-DNA, and thus also the DNA molecule inserted therein, are installed in the chromosomal DNA of the cell in a stable manner so that the genes of the T-DNA will be expressed in the plant.

There exist numerous known, efficient transfection mechanisms for different host systems. Some examples are electroporation, the calcium phosphate method, microinjection, liposome method.

Subsequently, the transfected cells are selected, e.g. on the basis of antibiotic resistences for which the vector comprises genes, or other marker genes. Then the transfected cell lines are amplified, either in small amounts, e.g. in Petri dishes, or in large amounts, e.g. in fermentors. Furthermore, plants have a particular characteristic, i.e. they are capable to re-develop from one (transfected) cell or from a protoplast, respectively, to a complete plant which can be grown.

Depending on the vector used, processes will occur in the host so that the enzyme expression will be suppressed or completely blocked:

If the vector comprising the DNA molecule with the deletion, insertion or substitution mutation is transfected, a homologous recombination will occur: the mutant DNA molecule will recognize the identical sequence in the genome of the host cell despite its

mutation and will be inserted exactly on that place so that a "knock-out gene" is formed. In this manner, a mutation is introduced into the gene for the GlcNAc-α1,3-fucosyl transferase which is capable of inhibiting the faultless expression of the GlcNAc-α1,3-fucosyl transferase. As has been explained above, with this technique it is important that the mutation suffices to block the expression of the active protein. After selection and amplification, the gene may be sequenced as an additional check so as to determine the success of the homologous recombination or the degree of mutation, respectively.

If the vector comprising the DNA molecule coding for a ribozyme is transfected, the active ribozyme will be expressed in the host cell. The ribozyme complexes the complementary mRNA sequence of the GlcNAc-α1,3-fucosyl transferase at least at a certain site, cleaves this site, and in this manner it can inhibit the translation of the enzyme. In this host cell as well as in cell lines, or optionally, plant, respectively, derived therefrom, GlcNAc-α1,3-fucosyl transferase will not be expressed. In case the vector comprises the inventive DNA molecule in sense or inverse direction to the promoter, a sense or antisense-mRNA will be expressed in the transfected cell (or plant, respectively). The antisense mRNA is complementary at least to a part of the mRNA sequence of the GlcNAc-α1,3-fucosyl transferase and may likewise inhibit translation of the enzyme. As an example of a method of suppressing the expression of a gene by antisense technique, reference is made to the publication by Smith et al., 1990, Mol. Gen. Genet. 224:477-481, wherein in this publication the expression of a gene involved in the maturing process of tomatoes is inhibited.

In all the systems, expression of the GlcNAc-α1,3-fucosyl transferase is at least suppressed, preferably even completely blocked. The degree of the disturbance of the gene expression will depend on the degree of complexing, homologous recombination, on possible subsequent coincidental mutations and on other processes in the region of the genome. The transfected cells are checked for GlcNac-α1,3-fucosyl transferase activity and selected.

Moreover, it is possible to still further increase the above-described suppression of the expression of the α1,3-fucosyl transferase by introducing into the host a vector comprising a

gene coding for a mammalian protein, e.g. β1,4-galactosyl transferase, in addition to the insertion of an above-described vector. Fucosylation may be reduced by the action of other mammalian enzymes, the combination of the inhibition of the expression of an active α1,3-fucosyl transferase by means of the inventive vector and by means of a mammalian enzyme vector being particularly efficient.

Any type of plant may be used for transfection, e.g. mung bean, tobacco plant, tomato and/or potato plant.
Another advantageous method of producing recombinant host cells, in particular plant or insect cells, or plants or insects, respectively, consists in that the DNA molecule comprising the mutation is inserted into the genome of the host cell, or plant or insect, respectively, in the place of the non-mutant homologous sequence (Schaefer et al., 1997, Plant J.; 11(6):1195-1206). This method thus does not function with a vector, but with a pure DNA molecule. The DNA molecule is inserted into the host e.g. by gene bombardment, microinjection or electroporation, to mention just three examples. As has already been explained, the DNA molecule binds to the homologous sequence in the genome of the host so that a homologous recombination and thus reception of the deletion, insertion or substitution mutation, respectively, will result in the genome: Expression of the GlcNAc-α1,3-fucosyl transferase can be suppressed or completely blocked, respectively.

A further aspect of the invention relates to plants or plant cells, respectively, as well as insect or insect cells, respectively, their GlcNAc-α1,3-fucosyl transferase activity being less than 50%, in particular less than 20%, particularly preferred 0%, of the GlcNAc-α1,3-fucosyl transferase activity occurring in natural plants or plant cells, respectively, and insects or insect cells, respectively. The advantage of these plants or plant cells, respectively, is that the glycoproteins produced by them do not comprise any or hardly comprise any α1,3-bound fucose. If products of these plants or insects, respectively, are taken up by human or vertebrate bodies, there will be no immune reaction to the α1,3-fucose epitope.

Preferably, recombinant plants or plant cells, respectively, are provided which have been prepared by one of the methods described above, their GlcNAc-α1,3-fucosyl transferase production

being suppresed or completely blocked, respectively.

The invention also relates to recombinant insects or insect cells, respectively, which have been prepared by one of the methods described above and whose GlcNAc-α1,3-fucosyl transferase production is suppressed or completely blocked, respectively. Also in this instance, no glycoproteins having α1,3-bound fucose residues are produced so that likewise no immune reaction to the α1,3-fucose epitope will occur.

The invention also relates to a PNA molecule comprising a base sequence complementary to the sequence of the DNA molecule according to the invention as well as partial sequences thereof. PNA (peptide nucleic acid) is a DNA-like sequence, the nucleobases being bound to a pseudo-peptide backbone. PNA generally hybridizes with complementary DNA-, RNA- or PNA-oligomers by Watson-Crick base pairing and helix formation. The peptide backbone ensures a greater resistance to enzymatic degradation. The PNA molecule thus is an improved antisense agent. Neither nucleases nor proteases are capable of attacking a PNA molecule. The stability of the PNA molecule, if bound to a complementary sequence, comprises a sufficient steric blocking of DNA and RNA polymerases, reverse transcriptase, telomerase and ribosomes. If the PNA molecule comprises the above-mentioned sequence, it will bind to the DNA or to a site of the DNA, respectively, which codes for GlcNAc-α1,3-fucosyl transferase and in this way is capable of inhibiting transcription of this enzyme. As it is neither transcribed nor translated, the PNA molecule will be prepared synthetically, e.g. by aid of the the t-Boc technique. Advantageously, a PNA molecule is provided which comprises a base sequence which corresponds to the sequence of the inventive DNA molecule as well as partial sequences thereof. This PNA molecule will complex the mRNA or a site of the mRNA of GlcNAc-α1,2-fucosyl transferase so that the translation of the enzyme will be inhibited. Similar arguments as set forth for the antisense RNA apply in this case. Thus, e.g., a particularly efficient complexing region is the translation start region or also the 5'-non-translated regions of mRNA.

A further aspect of the present invention relates to a method of preparing plants or insects, or cells, respectively, in particular plant or insect cells which comprise a blocked expression of the GlcNAc-α1,3-fucosyl transferase on transcription or

translation level, respectively, which is characterized in that inventive PNA molecules are inserted in the cells. To insert the PNA molecule or the PNA molecules, respectively, in the cell, again conventional methods, such as, e.g., electroporation or microinjection, are used. Particularly efficient is insertion if the PNA oligomers are bound to cell penetration peptides, e.g. transportan or pAntp (Pooga et al., 1998, Nature Biotechnology, 16; 857-861).

The invention provides a method of preparing recombinant glycoproteins which is characterized in that the inventive, recombinant plants or plant cells, respectively, as well as recombinant insects or insect cells, respectively, whose GlcNAc-α1,3-fucosyl transferase production is suppressed or completely blocked, respectively, or plants or insects, or cells, respectively, in which the PNA molecules have been inserted according to the method of the invention, are transfected with the gene that expresses the glycoprotein so that the recombinant glycoproteins are expressed. In doing so, as has already been described above, vectors comprising genes for the desired proteins are transfected into the host or host cells, respectively, as has also already been described above. The transfected plant or insect cells will express the desired proteins, and they have no or hardly any α1,3-bound fucose. Thus, they do not trigger the immune reactions already mentioned above in the human or vertebrate body. Any proteins may be produced in these systems. Advantageously, a method of preparing recombinant human glycoproteins is provided which is characterized in that the recombinant plants or plant cells, respectively, as well as recombinant insects or insect cells, respectively, whose GlcNAc-α1,3-fucosyl transferase production is suppressed or completely blocked, or plants or insects, or cells, respectively, in which PNA molecules have been inserted according to the method of the invention, are transfected with the gene that expresses the glycoprotein so that the recombinant glycoproteins are expressed. By this method it becomes possible to produce human proteins in plants (plant cells) which, if taken up by the human body, do not trigger any immune reaction directed against α1,3-bound fucose residues. There, it is possible to utilize plant types for producing the recombinant glycoproteins which serve as food stuffs, e.g. banana, potato and/or tomato. The tissues of this plant comprise

the recombinant glycoprotein so that, e.g. by extraction of the recombinant glycoprotein from the tissue and subsequent administration, or directly by eating the plant tissue, respectively, the recombinant glycoprotein is taken up in the human body. Preferably, a method of preparing recombinant human glycoproteins for medical use is provided, wherein the inventive, recombinant plants or plant cells, respectively, as well as recombinant insects or insect cells, respectively, whose GlcNAc-α1,3-fucosyl transferase production is suppressed or completely blocked, respectively, or plants or insects, or cells, respectively, into which the PNA molecules have been inserted according to the method of the invention, are transfected with the gene that expresses the glycoprotein so that the recombinant glycoproteins are expressed. In doing so, any protein can be used which is of medical interest.

Moreover, the present invention relates to recombinant glycoproteins according to a method described above, wherein they have been prepared in plant or insect systems and wherein their peptide sequence comprises less than 50%, in particular less than 20%, particularly preferred 0%, of the α1,3-bound fucose residues occurring in proteins expressed in non-fucosyl transferase-reduced plant or insect systems. Naturally, glycoproteins which do not comprise α1,3-bound fucose residues are to be preferred. The amount of α1,3-bound fucose will depend on the degree of the above-described suppression of the GlcNAc-α1,3-fucosyl transferase.

Preferably, the invention relates to recombinant human glycoproteins which have been produced in plant or insect systems according to a method described above and whose peptide sequence comprises less than 50%, in particular less than 20%, particularly preferred 0%, of the α1,3-bound fucose residues occurring in the proteins expressed in non-fucosyl transferase-reduced plant or insect systems.

A particularly preferred embodiment relates to recombinant human glycoproteins for medical use which have been prepared in plant or insect systems according to a method described above and whose peptide sequence comprises less than 50%, in particular less than 20%, particularly preferred 0%, of the α1,3-bound fucose residues occurring in the proteins expressed in non-fucosyl transferase-reduced plant or insect systems.

The glycoproteins according to the invention may include other bound oligosaccharide units specific for plants or insects, respectively, whereby - in the case of human glycoproteins - they differ from these natural glycoproteins. Nevertheless, by the glycoproteins according to the invention, a slighter immune reaction or no immune reaction at all, respectively, is triggered in the human body, since, as has already been explained in the introductory portion of the specification, the α1,3-bound fucose residues are the main cause for the immune reactions or cross immune reaction, respectively, to plant and insect glycoproteins.

A further aspect comprises a pharmaceutical composition comprising the glycoproteins according to the invention. In addition to the glycoproteins of the invention, the pharmaceutical composition comprises further additions common for such compositions. These are, e.g., suitable diluting agents of various buffer contents (e.g. Tris-HCl, acetate, phosphate, pH and ionic strength, additives, such as tensides and solubilizers (e.g. Tween 80, Polysorbate 80), preservatives (e.g. Thimerosal, benzyl alcohol), adjuvants, antioxidants (e.g. ascorbic acid, sodium metabisulfite), emulsifiers, fillers (e.g. lactose, mannitol), covalent bonds of polymers, such as polyethylene glycol, to the protein, incorporation of the material in particulate compositions of polymeric compounds, such as polylactic acid, polyglycolic acid, etc. or in liposomes, auxiliary agents and/or carrier substances which are suitable in the respective treatment. Such compositions will influence the physical condition, stability, rate of in vivo liberation and rate of in vivo excretion of the glycoproteins of the invention.

The invention also provides a method of selecting DNA molecules which code for a GlcNAc-α1,3-fucosyl transferase, in a sample, wherein the labelled DNA molecules of the invention are admixed to the sample, which bind to the DNA molecules that code for a GlcNAc-α1,3-fucosyl transferase. The hybridized DNA molecules can be detected, quantitated and selected. For the sample to contain single strand DNA with which the labelled DNA molecules can hybridize, the sample is denatured, e.g. by heating. One possible way is to separate the DNA to be assayed, possibly after the addition of endonucleases, by gele electrophoresis on an agarose gel. After having been transferred to a membrane of nitrocellulose, the labelled DNA molecules according to the in-

vention are admixed which hybridize to the corresponding homologous DNA molecule ("Southern blotting").

Another possible way consists in finding homologous genes from other species by PCR-dependent methods using specific and/or degenerated primers, derived from the sequence of the DNA molecule according to the invention.

Preferably, the sample for the above-identified inventive method comprises genomic DNA of a plant or insect organism. By this method, a large number of plants and insects is assayed in a very rapid and efficient manner for the presence of the GlcNAc-α1,3-fucosyl transferase gene. In this manner, it is respectively possible to select plants and insects which do not comprise this gene, or to suppress or completely block, respectively, the expression of the GlcNAc-α1,3-fucosyl transferase in such plants and insects which comprise this gene, by an above-described method of the invention, so that subsequently they may be used for the transfection and production of (human) glycoproteins. The invention also relates to DNA molecules which code for a GlcNAc-α1,3-fucosyl transferase which have been selected according to the two last-mentioned methods and subsequently have been isolated from the sample. These molecules can be used for further assays. They can be sequenced and in turn can be used as DNA probes for finding GlcNAc-α1,3-fucosyl transferases. These - labelled - DNA molecules will function for organisms, which are related to the organisms from which they have been isolated, more efficiently as probes than the DNA molecules of the invention. A further aspect of the invention relates to a preparation of GlcNAc-α1,3-fucosyl transferase cloned according to the invention which comprises isoforms having pI values of between 6.0 and 9.0, in particular between 6.8 and 8.2. The pI values of a protein is that pH value at which its net charge is zero and is dependent on the amino acid sequence, the glycosylation pattern as well as on the spatial structure of the protein. The GlcNAc-α1,3-fucosyl transferase comprises at least 7 isoforms which have a pI value in this range. The reason for the various isoforms of the transferase are, e.g., different glycosylations as well as limited proteolysis. Tests have shown that mung bean seedlings of various plants have different relationships of the isozymes. The pI value of a protein can be determined by isoelectric focussing, which is known to the skilled artisan.

The main isoform of the enzyme has an apparent molecular weight of 54 kDa.

In particular, the preparation of the invention comprises isoforms having pI values of 6.8, 7.1 and 7.6.

The invention also relates to a method of preparing "plantified" carbohydrate units of human and other vertebrate glycoproteins, wherein fucose units as well as GlcNAc-α1,3-fucosyl transferase encoded by an above-described DNA molecule are admixed to a sample that comprises a carbohydrate unit or a glycoprotein, respectively, so that fucose in α1,3-position will be bound by the GlcNAc-α1,3-fucosyl transferase to the carbohydrate unit or to the glycoprotein, respectively. By the method according to the invention for cloning GlcNAc-α1,3-fucosyl transferase it is possible to produce large amounts of purified enzyme. To obtain a fully active transferase, suitable reaction conditions are provided. It has been shown that the transferase has a particularly high activity at a pH of approximately 7, if 2-(N-morpholino)-ethane sulfonic acid-HCl is used as the buffer. In the presence of bivalent cations, in particular Mn^{2+}, the activity of the recombinant transferase is enhanced. The carbohydrate unit is admixed to the sample either in unbound form or bound to a protein. The recombinant transferase is active for both forms.
The invention will be explained in more detail by way of the following examples and drawing figures to which, of course, it shall not be restricted.

In detail, in the drawings,

Figs. 1a and 1b show, as curves, the measured amounts of protein and the measured enzyme activity in the individual fractions of the eluate;

Fig. 2 shows an electrophoresis gel analysis of GlcNAc-α1,3-fucosyl transferase;

Fig. 3 shows the result of the isoelectric focussing and the measured transferase activity of the individual isoforms;

Fig. 4 shows the N-terminal sequences of 4 tryptic peptides 1-4 as well as the DNA sequence of three primers, S1, A2 and A3;

Figs. 5a and 5b show the cDNA sequence of α1,3-fucosyl transferase;

Figs. 6a and 6b show the amino acid sequence of α1,3-fucosyl transferase derived therefrom;

Fig. 7 is a schematic representation of the α1,3-fucosyl transfe-

rase as well as the hydrophobicity of the amino acid residues;
Fig. 8 shows a comparison of the conserved motifs of various fucosyl transferases;
Fig. 9 shows a comparison of the fucosyl transferase activity of insect cells transfected with the α1,3-fucosyl transferase gene with that of a negative control;
Figs. 10a and 10b show structures of different acceptors of the α1,3-fucosyl transferase;
Figs. 11 and 12 show mass spectra; and
Fig. 13 shows the result of a HPLC.

Example 1:

Isolation of the core-α1,3-fucosyl transferase

All the steps were carried out at 4°C. Mung bean seedlings were homogenized in a mixer, 0.75 volumes of extraction buffer being used per kg of beans. Subsequently, the homogenate was filtered through two layers of cotton fabric, and the filtrate was centrifuged for 40 min at 30000xg. The supernatant was discarded, and the pellet was extracted with solution buffer over night with continuous stirring. Subsequent centrifugation at 30000xg for 40 min yielded the triton extract.

The triton extract was purified as follows:

Step 1: The triton extract was applied to a microgranular diethyl amino ethyl cellulose anion exchanger DE52 cellulose column (5x28 cm) from Whatman, which previously had been calibrated with buffer A. The non-bound fraction was further treated in step 2.

Step 2: The sample was applied to an Affi-Gel Blue column (2,5x32) column calibrated with buffer A. After washing of the column whith this buffer, adsorbed protein was eluted with buffer A comprising 0.5 M NaCl.

Step 3: After dialysis of the eluate from step 2 against buffer B, it was applied to an S-Sepharose column calibrated with the same buffer. Bound protein was eluted with a linear gradient of from 0 to 0.5 M NaCl in buffer B. Fractions with GlcNAc-α1,3-fucosyl transferase were pooled and dialyzed against buffer C.

Step 4: The dialyzed sample was applied to a GnGn-Sepharose column calibrated with buffer C. The bound protein was eluted with buffer C comprising 1 M NaCl instead of $MnCl_2$.

Step 5: Subsequently, the enzyme was dialyzed against buffer

D and applied to a GDP-Hexanolamine-Sepharose column. After having washed the column with buffer D, the transferase was eluted by substituting $MgCl_2$ and NaCl with 0.5 mM GDP. Active fractions were pooled, dialyzed against 20 mM Tris-HCl buffer, pH 7.3, and lyophilized.

The enzymatic activity of the GlcNAc-α1,3-fucosyl transferase was determined by using GnGn peptide and GDP-L-[U-^{14}C]-fucose at substrate concentrations of 0.5 and 0.25 each, in the presence of 2-(N-morpholino)ethanesulfonic acid-HCl buffer, Triton X-100, $MnCl_2$, GlcNAc and AMP (according to Staudacher et al., 1998, Glycoconjugate J. 15, 355-360; Staudacher et al., 1991, Eur. J. Biochem. 199, 745-751).

Protein concentrations were determined by aid of the bicinchoninic acid method (Pierce) or, in the final steps of enzyme purification, by means of amino acid analysis (Altmann 1992, Anal. Biochem. 204, 215-219).

In Figs. 1a and 1b, the measured amounts of protein and the measured enzyme activity in the individual fractions of the eluate are illustrated as curves. Fig. 1a shows the above-described separation on the S-Sepharose column, Fig. 1b shows the separation on the GnGn-Sepharose column, the circle representing protein, the black, full circle representing GlcNAc-α1,3-fucosyl transferase, and the square illustrating N-acetyl-β-glucosaminidase. One U is defined as that amount of enzyme which transfers 1 mmol of fucose onto an acceptor per minute.

Table 1 shows the individual steps of transferase purification.

Table 1

Purification step	Total protein	Total activity	Specific activity	Purification factor	Yield
	mg	mU	mU/mg	-fold	%
Triton X-100 extract	91500	4846	0.05	1	100
DE52	43700	4750	0.10	2	98.0
Affigel Blue	180.5	4134	23	460	85.3
S-Sepharose	8.4	3251	390	7800	67.1
GnGn-Sepharose	0.13[1]	1044	8030	160000	21.5
GDP-Hexanolamine-Sepharose	0.02[1]	867	43350	867000	17.9

[1]determined by amino acid analysis

Extraction buffer:

0.5 mM Dithiothreitol
1 mM EDTA
0.5% Polyvinyl polypyrrolidone
0.25 M Sucrose
50 mM Tris-HCl buffer, pH 7.3

Solution buffer:

0.5 mM Dithiothreitol
1 mM EDTA
1.5% Triton X-100
50 mM Tris-HCl, pH 7.3

Buffer A:

25 mM Tris-HCl buffer, pH 7.3, comprising:
0.1% Triton X-100 and
0.02% NaN_3

Buffer B:

25 mM Na citrate buffer, pH 5.3, comprising:
0.1% Triton X-100 and
0.02% NaN_3

Buffer C:

25 mM Tris-HCl buffer, pH 7.3, comprising:
5 mM $MnCl_2$nd
0.02% NaN_3

Buffer D:

25 mM Tris-HCl, pH 7.3, comprising:
10 mM $MgCl_2$
0.1 M NaCl, and
0.02% NaN_3

Example 2:

SDS-PAGE and isoelectric focussing

An SDS-PAGE was carried out in a Biorad Mini-protean cell on gels with 12.5% acrylamide and 1% bisacrylamide. The gels were stained either with Coomassie Brilliant Blue R-250 or Silver. Isoelectric focussing of the fucosyl transferase was carried out on prefabricated gels having a pI range of between 6-9 (Servalyt precotes 6-9, Serva). The gels were stained with silver according to the producer's protocol. For the two-dimensional electrophoresis, lanes were cut out of the focussing gel, treated with S-alkylating reagents and SDS and subjected to an SDS-PAGE, as described above.

Fig. 2 shows the illustration of an electrophoresis gel of GlcNAc-α1,3-fucosyl transferase, the two-dimensional electrophoresis being indicated on the left-hand side, and the one-dimensional SDS-PAGE being illustrated on the right-hand side. The lane denoted by A is a standard, the lane denoted by B is the GlcNAc-α1,3-fucosyl transferase from the GnGn-Sepharose column, and the lane denoted by C is the "purified" GlcNAc-α1,3-fucosyl transferase, i.e. the fraction of the GDP Hexanolamine Sepharose column. The two bands at 54 and 56 kDa represent isoforms of the transferase.

Fig. 3 shows the result of the isoelectric focussing. Lane A was stained with silver, on lane B, the activity of the transferase isoforms was tested. The activity is indicated as % fucose which had been transferred from GDP-fucose onto the substrate.

Example 3:

Peptide sequencing

For sequencing of the protein, bands were cut out of the Coomassie-stained SDS-Polyacrylamide gel, carboxyamido-methylated and cleaved with trypsin according to Görg et al. 1988, Electrophoresis, 9, 681-692. The tryptic peptides were separated with the reverse phase HPLC on a 1.0x250 mm Vydac C18 at 40°C at a flow rate of 0.05 ml/min, wherein a HP 1100 apparatus (Hewlett-

Packard) was used. The isolated peptides were separated with a Hewlett-Packard G1005 A Protein Sequencing System according to the producer's protocol. Furthermore, the peptide mixture was analyzed by Ingel digestion with MALDI-TOF MS (see below). Fig. 4 shows the N-terminal sequences of 4 tryptic peptides 1-4 (SEQ ID NO: 5-8). Departing from the first three peptides, primers S1, A2 and A3 were prepared (SEQ ID NO: 9-11).

Example 4:

RT-PCR and cDNA cloning

The entire RNA was isolated from a 3-day-old mung bean hypocotyl, wherein the SV Total RNA Isolating System of Promega was used. To prepare the first strand cDNA, the entire RNA was incubated for 1 h at 48°C with AMV reverse transcriptase and oligo(dT) primers, wherein the Reverse Transcription System of Promega was used. The first strand cDNA was subjected to a PCR, wherein a combination of sense and antisense primers was used:

To 10 µl of the reverse transcription reaction mixture, the following was added:

50 µl with 0.1 mmol of each primer, 0.1 mM dNTPs, 2 mM $MgCl_2$, 10 mM Tris-HCl buffer, pH 9,0, 50 mM KCl and 0.1% Triton X-100. After a first denaturing step at 95°C for 2 min, 40 cycles of 1 min at 95°C, 1 min at 49°C and 2 min at 72°C were passed. The last extension step was carried out at 72°C for 8 min. PCR products were subcloned into the pCR2.1 vector, with the TA Cloning Kit of Invitrogen being used, and sequenced. The products of this PCR were two DNA fragments with lengths of 744 bp and 780 bp, both DNA fragments having the same 5'-end (cf. also Fig. 7). Starting from these two DNA fragments, the missing 5' and 3' regions of the cDNA were obtained by 5' and 3' rapid amplification of cDNA ends (RACE), wherein the RACE Kit of Gibco-BRL was used. As the antisense primer, the universal amplification primer of the kit, and as the sense primer, either 5'-CTGGAACTGTCCCTGTGGTT-3' (SEQ ID NO: 12) or 5'- AGTGCACTAGAGGGCCAGAA-3' (SEQ ID NO: 13) were used. As the sense primer, also the shortened anchor primer of the kit, and as the antisense primer, 5'-TTCGAGCACCA-CAATTGGAAAT-3' (SEQ ID NO: 14) or 5'-GAATGCAAAGACGGCACGATGAAT-3' (SEQ ID NO: 15) were used.

The PCR was carried out with an annealing temperature of 55°C and under the above-described conditions. The 5' and 3' RACE products were subcloned into the pCR2.1 vector and sequenced: The se-

quences of the subcloned fragments were sequenced by means of the didesoxynucleotide method (ABI PRISM Dye Terminator Cycle Sequencing Ready reaction Kit and ABI PRISM 310 Genetic analyser (Perkin Elmer)). T7 and M13 forward primers were used for the sequencing of the products cloned into vector pCR2.1. Both strands of the coding region were sequenced by the Vienna VBC Genomics-Sequencing Service, infrared-labelled primers (IRD700 and IRD800) and an LI-COR Long Read IR 4200 Sequencer (Lincoln, NE) being used.

Figs. 5a and 5b show the entire cDNA which has a size of 2198 bp and an open reading frame of 1530 bp (SEQ ID NO: 1). The open reading frame (start codon at base pairs 211-213, stop codon at base pairs 1740-1743) codes for a protein of 510 amino acids having a molecular weight of 56.8 kDA and a theoretical pI value of 7.51.

Figs. 6a and 6b show the cDNA-derived amino acid sequence of the GlcNAc-α1,3-fucosyl transferase (SEQ ID NO: 2). Sites for the asparagine-bound glycosylation are at Asn346 and Asn429.

In Fig. 7, the schematic GlcNAc-α1,3-fucosyl transferase-cDNA (top) and the derived hydrophobicity index of the encoded protein (bottom) are illustrated, a positive hydrophobicity index meaning an increased hydrophobicity. Therebetween, the sizes of the two above-indicated PCR products are shown in relationship to the complete cDNA. The coding region is illustrated by the beam, "C" coding for the postulated cytoplasmatic region, T for the postulated transmembrane region, and G for the postulated Golgi lumen catalytic region of transferase. The analysis of the DNA sequence by "TMpred" (from EMBnet, Switzerland) gave an assumed transmembrane region between Asn36 and Gly54. The C-terminal region of the enzyme probably comprises the catalytic region and consequently should point into the lumen of the Golgi apparatus. According to this, this transferase seems to be a type II transmembrane protein like all the hitherto analyzed glycosyl transferases which are involved in glycoprotein biosynthesis (Joziasse, 1992, Glycobiology 2, 271-277). The gray regions represent the four tryptic peptides, the hexagons represent the potential N-glycosylation sites. A°BLASTP search in all data banks accesible via NCBI showed a similarity between the GlcNAc-α1,3-fucosyl transferase and other α1,3/4-fucosyl transferases, e.g. human fucosyl transferase VI. At 18-21% (examined by SIM-LALN-

VIEW, Expase, Switzerland), the total similarity was beyond any significance. Nevertheless, a sequence range of 35 amino acids (SEQ ID NO: 4) shows a strikingly high homology to other α1,3/4-fucosyl transferases (Fig. 8). This sequence region is located between Glu267 and Pro301 of SEQ ID NO: 2.

Example 5:

Expression of recombinant GlcNAc-α1,3-fucosyl transferase in insect cells

The encoding region of the assumed GlcNAc-α1,3-fucosyl transferase including cytoplasmatic and transmembrane region was amplified with the forward primer 5'-CGGCGGATCCGCAATTGAATGATG-3' (SEQ ID NO: 16) and reverse primer 5'-CCGGCTGCAGTACCATTTAGCGCAT-3' (SEQ ID NO: 17) by means of the Expand High Fidelity PCR System of Boehringer Mannheim. The PCR product was double-digested with PstI and BamHI and subcloned in alkaline phosphatase-treated baculovirus transfer vector pVL1393 which previously had been digested with PstI and BamHI. To ensure a homologous recombination, the transfer vector was co-transfected with Baculo Gold viral DNA (PharMingen, Sand Diego, CA) in Sf9 insect cells in IPL-41 Medium with lipofectin. After an incubation of 5 days at 27°C, various volumes of the supernatant with the recombinant virus were used for infecting the Sf21 insect cells. After an incubation of 4 days at 27°C in IPL-41 Medium with 5% FCS, the Sf1 cells were harvested and washed 2x with phosphate-buffered saline solution. The cells were resuspended in 25 mM Tris HCl buffer, pH 7.4, with 2% Triton X-100 and broken up by sonication on ice.

Example 6:

Assay for GlcNAc-α1,3-fucosyl transferase activity

The homogenate and the cell supernatant were assayed for GlcNAc-α1,3-fucosyl transferase. Blind samples were carried out with recombinant baculovirus which codes for the tobacco-GlcNAc-transferase I (Strasser et al., 1999, Glycobiology, in the process of printing).

Fig. 9 shows the measured enzyme activity of the recombinant GlcNAc-α1,3-fucosyl transferase as well as of the negative control. At best, the enzyme activity of the cotransfected cells and their supernatant was 30x higher than that of the negative controls. This endogenous activity which is measurable in the absence of the recombinant transferase, substantially comes from the insect-α1,6-fucosyl transferase and only a low percentage

thereof comes from the GlcNAc-α1,3-fucosyl transferase. Accordingly, the increase in the GlcNAc-α1,3-fucosyl transferase coming from the recombinant baculoviruses is far more than the 100-fold. The enzyme exhibited a broad maximum activity around a pH of 7.0, if the activity was measured in 2-(N-morpholino)-ethanesulfonic acid-HCl buffer. As is apparent in Table 2, the addition of bivalent cations, in particular Mn^{2+}, enhances the activity of the recombinant transferase.

Table 2

Additive (conc. 10 mM)	Relative Activity (Acceptor: GnGn-peptide)
	%
none	21
EDTA	18
$MnCl_2$	100
$CaCl_2$	82
$MgCl_2$	52
$CdCl_2$	44
$CoCl_2$	35
$CuCl_2$	3
$NiCl_2$	24
$ZnCl_2$	0.6

Table 3 shows that among the acceptors used, the GnGn-peptide exhibits the highest incorporation rates under standard test conditions, followed closely by GnGnF^{6}eptide and M5Gn-Asn. A transfer to the MM peptide could not be found, which MM peptide does not comprise the reducing GlcNAc-end at the 3-bound mannose. This structure seems to be necessary for the core fucosyl transferase. The recombinant transferase, moreover, was inactive relative to the acceptors commonly used, the α,3/4-fucosyl transferases used for determining the blood groups, which transfer the fucose to GlcNAc at the non-reducing ends of oligosaccharides. The apparent K_m-values for the acceptor substrate GnGn peptide, GnGnF^{6}peptide, M5Gn-Asn, and for the donor substrate GDP-fucose, were assessed to be 0.19, 0.13, 0.23 and 0.11, respectively. The structures of the molecules are illustrated in Figs. 10a and 10b.

Table 3

Acceptor Substrate	Rel. Activity	K_m-Value
	%	mM
GnGn-peptide	100	0.19
GnGnF^6-peptide	87	0.13
M5Gn-Asn	71	0.23
MM-peptide	0	
Galβ-4GlcNAc	0	
Galβ1-3GlcNAc	0	
Galβ1-3GlcNAcβ1-3Galβ1-4Glc	0	

Example 7:

Mass spectrometry of the fucosyl transferase product

Dabsylated GnGn hexapeptide (2 nmol) was incubated with the insect cell homogenate comprising the recombinant GlcNAc-α,3-fucosyl transferase (0.08 mU) in the presence of non-radioactive GDP-L-fucose (10 nmol), 2 (N-morpholino)-ethanesulfonic acid-HCl buffer, Triton X-100, $MnCl_2$, GlcNAc and AMP. A negative control was carried out with a homogenate of the infected insect cells for the blind samples. The samples were incubated for 16 h at 37°C and analyzed by means of MALDI TOF mass spectrometry. Mass spectrometry was performed on a DYNAMO (Therrmo BioAnalysis, Santa Fe, NM), a MALDI-TOF MS which is capable of dynamic extraction (synonym for late extraction). Two types of sample matrix preparations were used: peptides and dabsylated glycopeptides were dissolved in 5% formic acid, and aliquots were applied to the target, air-dried, and covered with 1% α-cyano-4-hydroxy cinnamic acid. Pyridyl-aminated glycans, reduced oligosaccharides and non-derivatized glycopeptides were diluted with water, applied to the target and air-dried. After addition of 2% 2.5-dihydroxy benzoic acid, the samples were immediately dried by applying a vacuum.

Fig. 11 shows the mass spectrum of these samples, A being the negative control: The main peak (S) shows the Dabsyl-Val-Gly-Glu-($GlcNAc_4Man_3$)Asn-Arg-Thr substrate, the calculated [M+H]$^+$ value being 2262.3. This substrate also appears as sodium addi-

tion product and as smaller ion which has been formed by fragmentation of the Azo function of the Dabsyl group, at (S*). A small product amount (P, $[M+H]^+$ = 2408.4) is a consequence of the endogenous α1,6-fucosyl transferase. The peak at m/z = 2424.0 shows the incomplete de-galactosylation of the substrate. The mass spectrum B shows the sample with recombinant α1,3-fucosyl transferase. The main peak (P) represents the fucosylated product, (P*) its fragmented ion.

In addition, aliquots of both samples were mixed with each other so as to obtain similar concentrations of substrate and product (sample A). This mixture was diluted with 0.1 M ammonium acetate, pH 4.0, comprising 10 mU of N-glycosidase A (sample B), or with 50 mM Tris/HCl, pH 8.5, comprising 100 mU (1 U hydrolyses 1 mmol of substrate per min) of N-glycosidase F (sample C). After 2 and 20 h, small aliquots of these mixtures were taken and analyzed by means of MALDI-TOF MS.

In Fig. 12, the three mass spectra of samples A, B and C are illustrated. The undigested sample A shows two main peaks: the substrate at 2261.4 m/z, and the fucosylated product at 2407.7 m/z. The middle curve shows the mass spectrum of sample B, treated with N-glycosidase A, which hydrolyses both glycopeptides. The peak at 963.32 constitutes the deglycosylated product. The lower curve shows the mass spectrum of sample C. The N-glycosidase F is not able to hydrolyse α1,3-fucosylated substrates, so that the spectrum has the peak at 2406.7 m/z of the fucosylated product, whereas the peak of the hydrolysed substrate appears at 963.08 m/z.

Example 8:

HPLC-analysis of the pyridyl-aminated fucosyl transferase product

The two above-described samples (fucosylated product and negative control) were digested with N-glycosidase A. The oligosaccharides obtained were pyridyl-aminated and analysed by means of reverse phase HPLC (Wilson et al., 1998, glycobiology 8, 651-661; Kubelka et al., 1994, Arch. Biochem. Giophys. 308, 148-157; Hase et al., 1984, J. Biochem. 95, 197-203).

In Fig. 13, the top diagram B represents the negative control, wherein in addition to the residual substrate (GnGn-peptide) α 1,6-fucosylated product is visible. A has a peak at a substan-

tially shorter retention time, which is specific of reducing fucose bound to GlcNAc-α1,3.
In the bottom diagram, the isolated transferase product prior to (curve A) and following (curve B) digestion by N-acetyl-βglucosaminidase was compared with MMF[3] honeybee phospholipase A_2 (curve C).

SEQUENCE LISTING

<110> Altmann Dr., Friedrich

<120> alpha 1,3-fucosyltransferase R35063

<130> fucosyltransferase gene

<140>
<141>

<160> 17

<170> PatentIn Ver. 2.1

<210> 1
<211> 2198
<212> DNA
<213> plant

<400> 1
actaactcaa acgctgcatt ttcttttttc tttcagggaa ccatccaccc ataacaacaa 60
aaaaaacaac agcaagctgt gtttttttta tcgttctttt tctttaaaca agcaccccca 120
tcatggaatc gtgctcataa cgccaaaatt ttccatttcc ctttgatttt tagtttattt 180
tgcggaattg gcagttgggg gcgcaattga atgatgggtc tgttgacgaa tcttcgaggc 240
tcgagaacag atggtgccca acaagacagc ttacccgttt tggctccggg aggcaaccca 300
aagaggaaat ggagcaatct aatgcctctt gttgttgccc ttgtggtcat cgcggagatc 360
gcgtttctgg gtaggttgga tatggccaaa aacgccgcca tggttgactc cctcgctgac 420
ttcttctacc gctctcgagc ggtcgttgaa ggtgacgatt tggggttggg tttggtggct 480
tctgatcgga attctgaatc gtatagttgt gaggaatggt tggagaggga ggatgctgtc 540
acgtattcga ggggcttttc caaagagcct atttttgttt ctggagctga tcaggagtgg 600
aagtcgtgtt cggttggatg taaatttggg tttagtgggg atagaaagcc agatgccgca 660
tttgggttac ctcaaccaag tggaacagct agcattctgc gatcaatgga atcagcagaa 720
tactatgctg agaacaatat tgccatggca agacggaggg gatataacat cgtaatgaca 780
accagtctat cttcggatgt tcctgttgga tatttttcat gggctgagta tgatatgatg 840
gcaccagtgc agccgaaaac tgaagctgct cttgcagctg ctttcatttc caattgtggt 900
gctcgaaatt tccggttgca agctcttgag gcccttgaaa aatcaaacat caaaattgat 960
tcttatggtg gttgtcacag gaaccgtgat ggaagagtga acaaagtgga agccctgaag 1020
cactacaaat ttagcttagc gtttgaaaat tcgaatgagg aagattatgt aactgaaaaa 1080
ttcttccaat cccttgttgc tggaactgtc cctgtggttg ttggtgctcc aaatattcag 1140
gactttgctc cttctcctgg ttcaatttta catattaaag agatagagga tgttgagtct 1200
gttgcaaaga ccatgagata tctagcagaa aatcccgaag catataatca atcattgagg 1260
tggaagtatg agggtccatc tgactccttc aaggcccttg tggatatggc agctgtgcat 1320
tcatcgtgcc gtctttgcat tcacttggcc acagtgagta gagagaagga agaaaataat 1380
ccaagcctta agagacgtcc ttgcaagtgc actagagggc cagaaaccgt atatcatatc 1440
tatgtcagag aaaggggaag gtttgagatg gagtccattt acctgaggtc tagcaattta 1500
actctgaatg ctgtgaaggc tgctgttgtt ttgaagttca catccctgaa tcttgtgcct 1560
gtatggaaga ctgaaaggcc tgaagttata agaggggga gtgctttaaa actctacaaa 1620

atatacccaa ttggcttgac acagagacaa gctctttata ccttcagctt caaaggtgat 1680
gctgatttca ggagtcactt ggagaacaat ccttgtgcca agtttgaagt cattttgtg 1740
tagcatgcgc taaatggtac ctctgctcta cctgaattag cttcacttag ctgagcacta 1800
gctagagttt taggaatgag tatggcagtg aatatggcat ggctttattt atgcctagtt 1860
tcttggccaa ctcattgatg ttttgtataa gacatcacac tttaatttta aacttgtttc 1920
tgtagaagtg caaatccata tttaatgctt agttttagtg ctcttatctg atcatctaga 1980
agtcacagtt cttgtatatt gtgagtgaaa actgaaatct aatagaagga tcagatgttt 2040
cactcaagac acattattac ttcatgttgt tttgatgatc tcgagctttt ttagtgtctg 2100
gaactgtccc tgtggtttga gcacctgtta ttgcttcagt gttactgtcc agtggttatc 2160
gtttttgacc tctaaaaaaa aaaaaaaaaa aaaaaaaa 2198

<210> 2
<211> 510
<212> PRT
<213> plant

<400> 2

Met Met Gly Leu Leu Thr Asn Leu Arg Gly Ser Arg Thr Asp Gly Ala
1 5 10 15

Gln Gln Asp Ser Leu Pro Val Leu Ala Pro Gly Gly Asn Pro Lys Arg
20 25 30

Lys Trp Ser Asn Leu Met Pro Leu Val Val Ala Leu Val Val Ile Ala
35 40 45

Glu Ile Ala Phe Leu Gly Arg Leu Asp Met Ala Lys Asn Ala Ala Met
50 55 60

Val Asp Ser Leu Ala Asp Phe Phe Tyr Arg Ser Arg Ala Val Val Glu
65 70 75 80

Gly Asp Asp Leu Gly Leu Gly Leu Val Ala Ser Asp Arg Asn Ser Glu
85 90 95

Ser Tyr Ser Cys Glu Glu Trp Leu Glu Arg Glu Asp Ala Val Thr Tyr
100 105 110

Ser Arg Gly Phe Ser Lys Glu Pro Ile Phe Val Ser Gly Ala Asp Gln
115 120 125

Glu Trp Lys Ser Cys Ser Val Gly Cys Lys Phe Gly Phe Ser Gly Asp
130 135 140

Arg Lys Pro Asp Ala Ala Phe Gly Leu Pro Gln Pro Ser Gly Thr Ala
145 150 155 160

Ser Ile Leu Arg Ser Met Glu Ser Ala Glu Tyr Tyr Ala Glu Asn Asn

```
                  165                 170                 175

Ile Ala Met Ala Arg Arg Arg Gly Tyr Asn Ile Val Met Thr Thr Ser
             180                 185                 190

Leu Ser Ser Asp Val Pro Val Gly Tyr Phe Ser Trp Ala Glu Tyr Asp
         195                 200                 205

Met Met Ala Pro Val Gln Pro Lys Thr Glu Ala Ala Leu Ala Ala Ala
     210                 215                 220

Phe Ile Ser Asn Cys Gly Ala Arg Asn Phe Arg Leu Gln Ala Leu Glu
225                 230                 235                 240

Ala Leu Glu Lys Ser Asn Ile Lys Ile Asp Ser Tyr Gly Gly Cys His
                  245                 250                 255

Arg Asn Arg Asp Gly Arg Val Asn Lys Val Glu Ala Leu Lys His Tyr
             260                 265                 270

Lys Phe Ser Leu Ala Phe Glu Asn Ser Asn Glu Glu Asp Tyr Val Thr
         275                 280                 285

Glu Lys Phe Phe Gln Ser Leu Val Ala Gly Thr Val Pro Val Val Val
     290                 295                 300

Gly Ala Pro Asn Ile Gln Asp Phe Ala Pro Ser Pro Gly Ser Ile Leu
305                 310                 315                 320

His Ile Lys Glu Ile Glu Asp Val Glu Ser Val Ala Lys Thr Met Arg
                  325                 330                 335

Tyr Leu Ala Glu Asn Pro Glu Ala Tyr Asn Gln Ser Leu Arg Trp Lys
             340                 345                 350

Tyr Glu Gly Pro Ser Asp Ser Phe Lys Ala Leu Val Asp Met Ala Ala
         355                 360                 365

Val His Ser Ser Cys Arg Leu Cys Ile His Leu Ala Thr Val Ser Arg
     370                 375                 380

Glu Lys Glu Glu Asn Asn Pro Ser Leu Lys Arg Arg Pro Cys Lys Cys
385                 390                 395                 400

Thr Arg Gly Pro Glu Thr Val Tyr His Ile Tyr Val Arg Glu Arg Gly
                  405                 410                 415

Arg Phe Glu Met Glu Ser Ile Tyr Leu Arg Ser Ser Asn Leu Thr Leu
```

420 425 430

Asn Ala Val Lys Ala Ala Val Val Leu Lys Phe Thr Ser Leu Asn Leu
435 440 445

Val Pro Val Trp Lys Thr Glu Arg Pro Glu Val Ile Arg Gly Gly Ser
450 455 460

Ala Leu Lys Leu Tyr Lys Ile Tyr Pro Ile Gly Leu Thr Gln Arg Gln
465 470 475 480

Ala Leu Tyr Thr Phe Ser Phe Lys Gly Asp Ala Asp Phe Arg Ser His
485 490 495

Leu Glu Asn Asn Pro Cys Ala Lys Phe Glu Val Ile Phe Val
500 505 510

<210> 3
<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:cDNA

<400> 3
gaagccctga agcactacaa atttagctta gcgtttgaaa attcgaatga ggaagattat 60
gtaactgaaa aattcttcca atcccttgtt gctggaactg tccct 105

<210> 4
<211> 35
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:peptide

<400> 4
Glu Ala Leu Lys His Tyr Lys Phe Ser Leu Ala Phe Glu Asn Ser Asn
1 5 10 15

Glu Glu Asp Tyr Val Thr Glu Lys Phe Phe Gln Ser Leu Val Ala Gly
20 25 30

Thr Val Pro
35

<210> 5
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:peptide

<400> 5
Lys Pro Asp Ala Xaa Phe Gly Leu Pro Gln Pro Ser Thr Ala Ser
1 5 10 15

<210> 6
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:peptide

<400> 6
Pro Glu Thr Val Tyr His Ile Tyr Val Arg
1 5 10

<210> 7
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:peptide

<400> 7
Met Glu Ser Ala Glu Tyr Tyr Ala Glu Asn Asn Ile Ala
1 5 10

<210> 8
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:peptide

<400> 8
Gly Arg Phe Glu Met Glu Ser Ile Tyr Leu
1 5 10

<210> 9
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 9
gcngartayt aygcngaraa yaayathgc 29

<210> 10
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 10
crtadatrtg rtanacngty tc 22

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 11
tadatnswyt ccatytcraa 20

<210> 12
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 12
ctggaactgt ccctgtggtt 20

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 13
agtgcactag agggccagaa 20

<210> 14
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 14
ttcgagcacc acaattggaa at 22

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 15
gaatgcaaag acggcacgat gaat 24

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 16
cggcggatcc gcaattgaat gatg 24

<210> 17
<211> 25
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:DNA

<400> 17
ccggctgcag taccatttag cgcat 25